

# A multilocus microsatellite phylogeny of the *Drosophila virilis* group

L Orsini, S Huttunen<sup>1</sup> and C Schlötterer

Institut für Tierzucht und Genetik, Josef Baumann Gasse 1, 1210 Wien, Austria

We used a set of 48 polymorphic microsatellites derived from *Drosophila virilis* to infer phylogenetic relationships in the *D. virilis* clade. Consistent with previous studies, *D. virilis* and *D. lummei* were the most basal species of the group. Within the *D. montana* phylad, the phylogenetic relationship could not be resolved. Special attention was given to the differentiation between *D. americana texana*, *D. americana americana* and *D. novamexicana*. Significant differences between these three groups were detected by  $F_{ST}$  analyses.

**Keywords:** *D. virilis* group; phylogeny; microsatellites

## Introduction

The *Drosophila virilis* group is one of the best-studied species groups in the subgenus *Drosophila*. The number of species included in the *D. virilis* group is object of disagreement among scientists, depending on the classification of *D. americana americana* and *D. americana texana* as two distinct species or alternatively as two chromosomal forms of the same species. Results obtained mainly from sequence polymorphism analysis support the hypothesis that *D. a. americana* and *D. a. texana* are two forms of the same species (McAllister and Charlesworth, 1999; McAllister and McVean, 2000; Vieira *et al.*, 2001; McAllister, 2002; McAllister, 2003; Vieira *et al.*, 2003). Phylogenetic analyses recognize the existence of 14 species in the *D. virilis* group, divided in the *D. virilis* and *D. montana* phylads (Spicer and Bell, 2002). Recently, the *D. virilis* group has been further subdivided into four lineages: the *D. virilis* phylad, and the *D. montana*, *D. littoralis* and *D. kanekoi* subphylads (Spicer, 1992; Spicer and Bell, 2002). The analyses of multiple individuals for nuclear genes, however, indicated considerable sequence divergence within species resulting in groupings not consistent with the presumed phylogeny (Hilton and Hey, 1996; Nurminsky *et al.*, 1996; Hilton and Hey, 1997). It has been argued that this is a reflection of a high proportion of shared ancestral polymorphism in the *D. virilis* group.

Recently, microsatellite polymorphism analysis was successfully used to infer the phylogeny of the *D. melanogaster* group, which also shares ancestral poly-

Similarly, a model-based clustering method for multilocus genotype data also provided strong support for the presence of three differentiated groups. This genome-wide differentiation between *D. americana texana* and *D. americana americana* contrasts with previous analyses based on DNA sequence data.

Heredity (2004) 93, 161–165, advance online publication, 9 June 2004; doi:10.1038/sj.hdy.6800487

morphism (Harr *et al.*, 1998; Harr and Schlötterer, 2004). Interestingly, the microsatellite-based phylogeny of the *D. melanogaster* group is identical to the one inferred by sequence analysis of the *Odysseus* gene. As this gene is involved in sterility of hybrid males produced by crossing *D. simulans* and *D. mauritiana* (Ting *et al.*, 1998, 2000), phylogenies of *Odysseus* and other genes involved in hybrid sterility are more likely to reflect the species phylogeny than other, randomly selected genes. In the absence of information about genes involved in hybrid sterility, a consensus phylogeny based on multiple neutrally evolving loci should also reflect the species phylogeny (Schlötterer, 2001).

We used 48 microsatellite loci isolated from *D. virilis* to infer the phylogenetic relationships in the *D. virilis* phylad. Special emphasis was given to the *D. virilis* subphylad encompassing *D. virilis*, *D. a. americana*, *D. a. texana*, *D. novamexicana* and *D. lummei*. While our analysis indicated that members of the *D. virilis* group are well separated from each other, only very low resolution was observed for the *D. montana* phylad.

## Material and methods

The list of species used in this work is shown in Table 1 in the online supplementary material. A total of 48 microsatellite loci isolated from *D. virilis* (Schlötterer, 2000; Huttunen and Schlötterer, 2002) were used for phylogenetic analyses. DNA was extracted from a single individual from each strain using a high salt extraction protocol (Miller *et al.*, 1988). Microsatellite analysis followed standard protocols (Schlötterer, 1998). In brief, PCR reactions were performed in a final volume of 10 µl containing 50–100 ng of genomic DNA, 1 µM of each primer (the forward one was end-labeled with <sup>32</sup>P), 200 µM dNTPs, 1.5 mM MgCl<sub>2</sub> and 1 U *Taq* Polymerase. In total, 30 PCR cycles were performed with the following profile: 1 min at 95°C, 30 s at 45–59°C (depend-

Correspondence: C Schlötterer, Institut für Tierzucht und Genetik, Josef Baumann Gasse 1, 1210 Wien, Austria.

E-mail: christian.schloetterer@vfu-wien.ac.at

<sup>1</sup>Current address: Department of Biology, University of Oulu, PO Box 3000, 90014, Oulu, Finland.

Received 25 October 2003; accepted 19 February 2004; published online 9 June 2004

ing on the locus) and 30 s at 72°C. An initial denaturing step of 5 min at 95°C and a terminal extension of 30 min at 72°C were applied. PCR products were separated on a 7% denaturing polyacrylamide gel (32% formamide, 5.6 M urea) at 90 W and visualized by autoradiography after 12–24 h. Allele sizes were determined running a 'PCR slippage ladder' and a known size standard adjacent to the samples (Schlötterer and Zangerl, 1999).

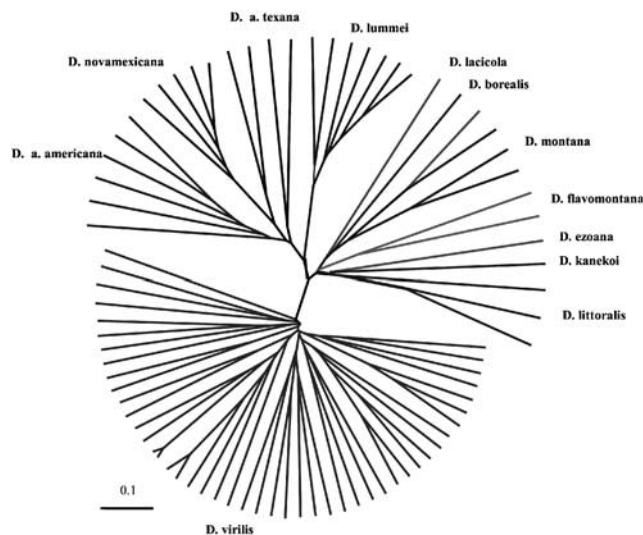
### Statistical analyses

For all analyses, we randomly selected one allele for each individual/locus to account for the inbreeding effect during the propagation of the isofemale lines. All genetic distances including the proportion of shared alleles, ( $\delta\mu$ )<sup>2</sup>, and  $F_{ST}$  were calculated with the Microsatellite Analyser Program (MSA) version 3.12 (Dieringer and Schlötterer, 2003). The genetic distance based on the proportion of shared allele was calculated as  $-\ln$  (proportion of shared allele). Standard statistical analyses were performed with the SPSS 10 program. UPGMA and Neighbor Joining (NJ) trees for species and populations were generated using PHYLIP 3.6 (Felsenstein, 1991). Phylogenetic trees were visualized using the Treeview 1.5 (Page, 1996).

The program STRUCTURE (Pritchard *et al.*, 2000a) was used to assign individuals to homogenous clusters (species) without consideration of the original species assignment. This program uses a Bayesian model-based clustering method for multilocus genotypes to simultaneously determine the most probable number of homogenous groups in a given data set and assigns individuals to one or more of them. The number of clusters is inferred by calculating the probability  $P(X|K)$  of the data, given a certain prior value of  $K$  (number of clusters) over a number of Monte Carlo Markov Chain (MCMC) iterations. The posterior probabilities  $P(K|X)$  can be calculated following Bayes' rule. The clusters are characterized by different allele frequencies and individuals are probabilistically assigned to one or more clusters, according to their allele distribution. The scores of individuals in the clusters correspond to the probability of ancestry in any one of them. In this study, we assumed prior values of  $K$  from 1 to 5. All calculations shown in this report are based on 1 000 000 iterations of the MCMC, following a 'burn-in' period of 50 000 iterations (the burn-in period is the first set of iterations of the MCMC that is dependent on the start configuration – the iterations are not incorporated in the final calculation of the posterior probability). We ran the program without incorporation of prior species information and allele frequencies were treated as uncorrelated.

### Results and discussion

Reflecting the public availability of strains, our data set was not balanced with respect to the number of individuals representing a given species. For some species, only a single line was used, while for others multiple lines were available. Therefore, the phylogenetic analysis was performed on the basis of individuals (taken from separate isofemale lines) rather than species. Both the UPGMA and the NJ trees based on the proportion of shared alleles (Bowcock *et al.*, 1994) indicate that isofemale lines from the same species tend grouping together. The phylogenetic trees obtained with UPGMA



**Figure 1** UPGMA tree of the whole set of individuals considered, using the proportion of shared alleles distance.

(Figure 1, Figure 2 supplementary material) and NJ (Figure 1 supplementary material) were very similar, but the UPGMA tree provides a better clustering according to species status. In the *D. virilis* phylad, *D. novamexicana* and *D. a. texana* are more closely related than *D. a. americana* and *D. a. texana*. *D. lummei* is well separated from these three species.

The phylogenetic relationship in the *D. montana* phylad is not resolved, despite that *D. littoralis*, *D. montana* and *D. flavomontana* isofemale lines group according to species status. The two *D. eozana* lines do not group together. One line clusters with *D. montana* while the other one groups with *D. kanekoi*. In the NJ tree, both individuals also group to different clades (Figure 2 supplementary material). While it would have been desirable to test whether this grouping is the consequence of insufficient statistical power, or has a biological meaning or is an artefact due to contamination of one of the lines, we could not address this question further as we had no alive flies after the analyses were completed.

One very interesting aspect of the microsatellite data is the low differentiation between species. In fact, the tree of individuals shown in Figure 1 resembles more an intraspecific tree of individuals rather than a tree containing different species. In the *D. melanogaster* group, microsatellite analysis based on the proportion of shared alleles separated the four species by long branches and high bootstrap (Harr *et al.*, 1998; Harr and Schlötterer, 2004). The low differentiation among species in the *D. virilis* phylad may have two reasons. First, as *D. virilis* microsatellites are longer than *D. melanogaster* (Schlötterer and Harr, 2000a), their mutation rate may be higher. Thus, alleles may be frequently identical by state but not by descent (homoplasy), rendering the proportion of shared alleles a problematic genetic distance. Alternatively, the species in the *D. virilis* group may share substantial ancestral variation or they are exposed to gene flow, which results in a low genetic differentiation. In order to discriminate between high mutation rate and shared ancestral alleles, we calculated pairwise genetic distances among species. Two genetic distance measure-

**Table 1** Pairwise genetic distances among species, calculated for the proportion of shared alleles (lower triangular matrix) and  $(\delta\mu)^2$  (upper triangular matrix)

	<i>D. ame</i>	<i>D. bor</i>	<i>D. ezo</i>	<i>D. flav</i>	<i>D. kan</i>	<i>D. lac</i>	<i>D. lit</i>	<i>D. lum</i>	<i>D. mon</i>	<i>D. nov</i>	<i>D. tex</i>	<i>D. vir</i>
<i>D. ame</i>	—	58.8	59.2	53.0	52.4	45.9	78.5	22.2	53.8	26.7	13.0	32.8
<i>D. bor</i>	2.4	—	26.3	21.8	47.7	32.9	52.3	39.9	20.9	72.1	45.4	62.4
<i>D. ezo</i>	2.9	1.9	—	34.5	34.7	35.8	26.4	53.2	18.2	89.0	52.9	54.5
<i>D. flav</i>	2.4	1.7	1.6	—	70.9	34.4	45.4	60.4	30.0	62.8	45.5	77.9
<i>D. kan</i>	2.8	2.4	2.2	2.51	—	36.7	69.5	55.3	50.3	77.4	44.4	58.9
<i>D. lac</i>	2.6	2.0	2.1	2.9	2.5	—	52.8	33.2	21.9	70.7	39.1	38.0
<i>D. lit</i>	2.4	2.4	1.6	2.2	2.4	2.5	—	73.3	32.6	117.7	55.7	94.9
<i>D. lum</i>	1.6	2.9	3.1	2.7	2.5	2.7	2.5	—	61.2	55.7	27.0	42.4
<i>D. mon</i>	2.6	1.6	1.2	2.0	2.6	1.7	1.6	2.1	—	55.8	38.9	70.5
<i>D. nov</i>	1.3	4.3	2.8	2.0	2.1	3.1	2.9	1.8	2.5	—	35.7	85.1
<i>D. tex</i>	1.3	2.7	2.3	2.3	2.6	3.2	2.6	2.0	2.3	1.2	—	36.7
<i>D. vir</i>	1.7	2.8	2.2	2.8	2.1	3.0	2.3	1.9	2.8	2.9	2.2	—

*D. ame* = *D. americana americana*; *D. bor* = *D. borealis*; *D. ezo* = *D. ezona*; *D. flav* = *D. flavomontana*; *D. kan* = *D. kanekoi*; *D. lac* = *D. laticola*; *D. lit* = *D. littoralis*; *D. lum* = *D. lummei*; *D. mon* = *D. montana*; *D. nov* = *D. novamexicana*; *D. tex* = *D. a. texana*; *D. vir* = *D. virilis*.

ments were used – the proportion of shared alleles and  $(\delta\mu)^2$ . As  $(\delta\mu)^2$  is linear with time and mutation rate (Goldstein *et al*, 1995), genetic distances based on  $(\delta\mu)^2$  should reflect true divergence times. The proportion of shared alleles, however, asymptotes for long divergence times and/or high mutation rates (Goldstein *et al*, 1995). If the low differentiation between species is caused by a high microsatellite mutation rate, then the pairwise distances based on  $(\delta\mu)^2$  and proportion of shared alleles should not be (or only loosely) correlated. Nevertheless, we found a highly significant correlation between them (Kendall's  $\tau = 0.411$ ,  $P < 0.001$ ). Also the coefficient of variation of all pairwise distances was similar for  $(\delta\mu)^2$  and proportion of shared alleles (0.42 *vs* 0.28). The slightly higher coefficient of variation observed for  $(\delta\mu)^2$  may indicate that this distance measurement provides more differentiation among the species studied (Table 1). Nevertheless, it has to be noted that  $(\delta\mu)^2$  is highly sensitive to indels occurring in the sequence flanking the microsatellite repeat. As such indels occur frequently (Colson and Goldstein, 1999), they may inflate  $(\delta\mu)^2$  based distances. Given that we failed to observe obvious differences between two distance measurements, of which one is highly sensitive to a high mutation rate, but the other ( $(\delta\mu)^2$ ) is not, we conclude that the low differentiation between the species in the *D. virilis* group is not primarily caused by a high microsatellite instability.

Apart from homoplasy (allele sharing generated by mutation rather by shared descent), the low differentiation among the species in the *D. virilis* group could have two reasons. Either there is still ongoing gene flow among the species or they maintained a large effective population size, which resulted in a large number of shared ancestral alleles. For *D. a. americana* and *D. a. texana*, these two hypotheses have been intensively discussed. A fusion between element B and the X chromosome in *D. a. americana* differentiates *D. a. americana* and *D. a. texana* (Throckmorton, 1982). Nevertheless, sequence analysis of several genes provided ambiguous results (Hilton and Hey, 1996; Nurminsky *et al*, 1996; Hilton and Hey, 1997; Vieira and Charlesworth, 2000; O'Grady *et al*, 2001). Sequence analysis data of *Alcohol dehydrogenase* (*Adh*) (McAllister and Charlesworth, 1999) and a phylogenetic study performed using the *Adh* gene (Nurminsky *et al*, 1996) revealed the

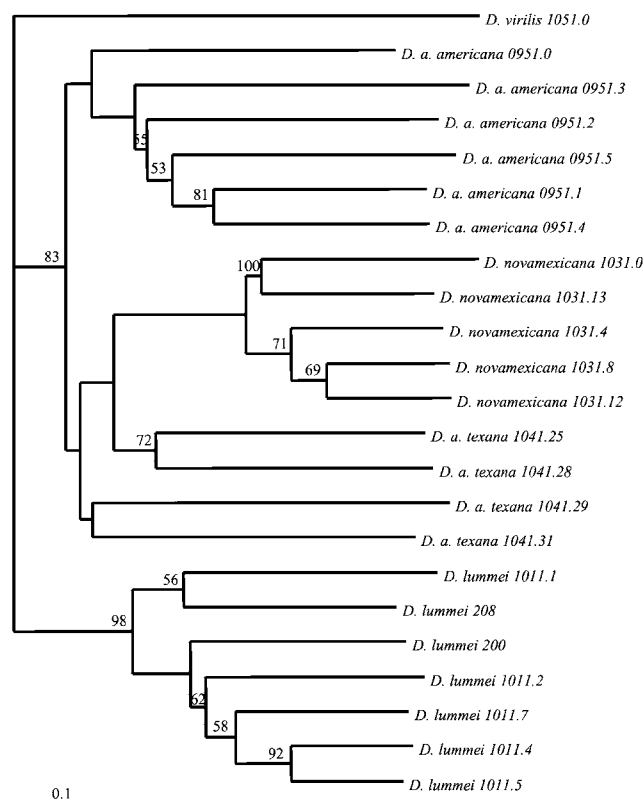
absence of divergence between *D. a. americana* and *D. a. texana*. Furthermore, the X-linked *per* gene (Hilton and Hey, 1996) and a population survey of RFLP variation on chromosome 4 failed to support genetic differentiation between *D. a. americana* and *D. a. texana* (McAllister, 2002). On the contrary, a highly significant differentiation between *D. a. americana* and *D. a. texana* was detected using microsatellites (Schlötterer, 2000) and the *fused1* locus (Vieira *et al*, 2001).

More recently, the absence of differentiation between *D. a. americana* and *D. a. texana* was assessed through the study of sequence polymorphism of four different genes (McAllister, 2003; Vieira *et al*, 2003).

We evaluated the statistical significance of the differentiation in the *D. virilis* phylad by generating pseudo-replicas of the microsatellite data by bootstrapping. To reduce the number of nodes, all *D. virilis* individuals were grouped into a single OTU, but for the other species of the *D. virilis* phylad each isofemale line was treated as separate OTU (Figure 2).

The monophyletic origin of *D. lummei* and *D. novamexicana* is strongly supported. This result is surprising as DNA sequence analysis of two genes in *D. novamexicana* indicated the presence of two diverged clusters clearly predating the speciation event (Hilton and Hey, 1996). The authors noted, however, that the assignment of individuals to one of the clades differed among the analyzed genes. As our data set is based on 48 microsatellite loci, such clusters formed by divergent ancestral alleles are very likely to be averaged out. Thus, the strong support for the monophyletic origin of *D. novamexicana* demonstrates the advantage of multilocus phylogenies.

The grouping of *D. a. texana* with *D. novamexicana* is only weakly supported. Furthermore, the  $(\delta\mu)^2$  distance grouped *D. a. texana* and *D. a. americana* together and *D. lummei* grouped closer to *D. a. texana* and *D. a. americana* than *D. novamexicana* (data not shown), a result not consistent with previous analyses (Hilton and Hey, 1996; O'Grady *et al*, 2001; Spicer and Bell, 2002). The statistical support for these groupings was very weak in a tree of individuals. Furthermore, other distance measurements – Nei's chord distance (Nei *et al*, 1983) and Cavalli Sforza and Edwards chord distance (Cavalli-Sforza and Edwards, 1967) – also supported the tree topology obtained by the proportion of shared alleles (data not shown). The



**Figure 2** NJ tree based on the proportion of shared allele of the *D. virilis* phylad. All *D. virilis* specimens are collapsed in a single OTU. Bootstrap values above 50% are shown.

results were essentially the same when UPGMA clustering algorithm was used rather than NJ (data not shown). The strong separation we observed between *D. a. americana* and *D. a. texana* could be due to the presence of high number of chromosomal inversions in *D. a. americana* that are present only at low frequency in *D. a. texana* (Hsu, 1952). The different frequency of inversions likely suppress recombination between *D. a. americana* and *D. a. texana* in a 600 Kb region around each inversion breakpoint and at the base of the X chromosome (Vieira et al., 2003). Therefore, recombination between chromosomes is prevented in a considerable proportion of the euchromatic genome. Microsatellites located in these genomic regions of reduced recombination between *D. a. americana* and *D. a. texana* are expected to show a more pronounced level of differentiation. The  $F_{ST}$  values per locus, however, indicated that 11 out of 43 loci (three loci did not amplify, vir44, vir63, v71-38 and two were monomorphic, vir35cs, vir32) were significantly different ( $P < 0.05$ ) between *D. a. americana* and *D. a. texana*. This result demonstrates that the phylogenetic differentiation between *D. a. americana* and *D. a. texana* is genome wide and not attributable to a low number of markers falling into the nonrecombining part of the chromosomes.

For further evidence for the separation of *D. a. texana* and *D. a. americana*, we used a model-based clustering method for multilocus genotype data to infer population/species structure and to assign individuals to groups (Pritchard et al., 2000a). In combination with Bayesian statistics, this method provides the most likely number of groups in the sample. Table 2 indicates that

**Table 2** Inferring the number of distinct groups in a sample consisting of *D. a. americana*, *D. a. texana* and *D. novamexicana* individuals

K	48 loci		42 loci <sup>a</sup>	
	$\ln P(X K)$	$P(K X)^b$	$\ln P(X K)$	$P(K X)^a$
1	-957.2	0.000	-807.4	0.000
2	-887.9	0.000	-760.1	0.000
3	-856.1	0.999	-727.6	1.000
4	-863.9	0.001	-737.5	0.000
5	-873.9	0.000	-746.8	0.000

<sup>a</sup>Excluding loci used in Schlötterer (2000).

<sup>b</sup>Assuming an uniform prior for K ( $K \in 1, 2, 3, 4, 5$ ).

**Table 3** Pairwise  $F_{ST}$  values (upper triangular matrix) and the corresponding P-values after Bonferroni correction (lower triangular matrix)

	<i>D. americana</i>	<i>D. texana</i>	<i>D. novamexicana</i>
<i>D. americana</i>	—	0.15	0.27
<i>D. texana</i>	0.011	—	0.25
<i>D. novamexicana</i>	0.007	0.027	—

the most likely number of groups is three. Individuals were assigned to each group according to their taxonomic status (ie: *D. a. americana*, *D. a. texana* and *D. novamexicana*). Only a single individual (*D. a. americana*, 0951.0) had a relatively high coefficient of coancestry with *D. a. texana* specimens ( $\approx 60\%$ ). This clear separation between *D. a. americana* and *D. a. texana* is fully consistent with a recent report using a small subset of loci mapping to Muller's element B (Schlötterer, 2000). As our data set contained the same loci used by Schlötterer (2000), we wanted to rule out that the signal of differentiation between *D. a. americana* and *D. a. texana* is limited to loci mapping to the fused neo-sex chromosomes. We repeated the analysis excluding all previously analyzed loci and obtained similar results (Table 2). Significant pairwise  $F_{ST}$  values between these three groups further substantiated our observation of a genome-wide differentiation between *D. a. americana*, *D. a. texana* and *D. novamexicana* (Table 3).

Previous studies also indicated that *D. a. texana* and *D. a. americana* are not the closest species pair (Hilton and Hey, 1996, 1997; Spicer and Bell, 2002). Nevertheless, it is surprising that while *D. a. americana* and *D. a. texana* share the fusion between Muller's element X and Y, *D. novamexicana* has the more ancestral karyotype lacking this fusion. A better sampling of *D. a. americana* and *D. a. texana* covering the distribution of both species is required to determine whether the significant differentiation between these species reflects geographic separation rather than different species status.

## Acknowledgements

We greatly acknowledge K Yoon from the former *Drosophila* species resource center, B Charlesworth, M Evgen'ev, A Hoikkala and J Aspi for providing flies. We thank members of the CS lab for helpful discussions and suggestions. This work was supported by EU and FWF grants to CS.

## References

- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**: 455–457.
- Cavalli-Sforza LL, Edwards AWF (1967). Phylogenetic analysis: models and estimation procedures. *Am J Hum Genet* **19**: 233–257.
- Colson I, Goldstein DB (1999). Evidence for complex mutations at microsatellite loci in *Drosophila*. *Genetics* **152**: 617–627.
- Dieringer D, Schlötterer C (2003). Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol Ecol Notes* **3**: 167–169.
- Felsenstein J (1991). Counting phylogenetic invariants in some simple cases. *J Theor Biol* **152**: 357–376.
- Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**: 463–471.
- Harr B, Schlötterer C (2004). Patterns of microsatellite variability in the *Drosophila melanogaster* complex. *Genetica* **120**: 71–77.
- Harr B, Weiss S, David JR, Brem G, Schlötterer C (1998). A microsatellite-based multilocus phylogeny of the *Drosophila melanogaster* species complex. *Curr Biol* **8**: 1183–1186.
- Hilton H, Hey J (1996). DNA sequence variation at the period locus reveals the history of species and speciation events in the *Drosophila virilis* group. *Genetics* **144**: 1015–1025.
- Hilton H, Hey J (1997). A multilocus view of speciation in the *Drosophila virilis* species group reveals complex histories and taxonomic conflicts. *Genet Res* **70**: 185–194.
- Hsu TC (1952). *Chromosomal Variation and Evolution in the Virilis Group of Drosophila*. University of Texas Publication: Texas.
- Huttunen S, Schlötterer C (2002). Isolation and characterization of microsatellites in *Drosophila virilis* and their cross species amplification in members of the *D. virilis* group. *Mol Ecol Notes* **2**: 593–597.
- McAllister BF (2002). Chromosomal and allelic variation in *Drosophila americana*: selective maintenance of a chromosomal cline. *Genome* **45**: 13–21.
- McAllister BF (2003). Sequence differentiation associated with an inversion on the neo-X chromosome of *Drosophila americana*. *Genetics* **165**: 1317–1328.
- McAllister BF, Charlesworth B (1999). Reduced sequence variability on the neo-Y chromosome of *Drosophila americana*. *Genetics* **153**: 221–233.
- McAllister BF, McVean GA (2000). Neutral evolution of the sex-determining gene transformer in *Drosophila*. *Genetics* **154**: 1711–1720.
- Miller SA, Dykes DD, Polesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acid Res* **16**: 1215.
- Nei M, Tajima F, Tatenno Y (1983). Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J Mol Evol* **19**: 153–170.
- Nurminsky DI, Moriyama EN, Lozovskaya ER, Hartl DL (1996). Molecular phylogeny and genome evolution in the *Drosophila virilis* species group: duplications of the alcohol dehydrogenase gene. *Mol Biol Evol* **13**: 132–149.
- O'Grady PM, Baker RH, Durando CM, Etges WJ, DeSalle R (2001). Polytene chromosomes as indicators of phylogeny in several species groups of *Drosophila*. *BMC Evol Biol* **1**: 6.
- Page RDM (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Appl Biosci* **12**: 357–358.
- Pritchard JK, Stephens M, Donnelly P (2000a). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Schlötterer C (1998). Microsatellites. In: Hoelzel AR (eds) *Molecular Genetic Analysis of Populations: a Practical Approach*. Oxford University Press: Oxford, pp 237–261.
- Schlötterer C (2000). Microsatellite analysis indicates genetic differentiation of the neo-sex chromosomes in *Drosophila americana americana*. *Heredity* **85**: 610–616.
- Schlötterer C (2001). Genealogical inference of closely related species based on microsatellites. *Genet Res* **78**: 209–212.
- Schlötterer C, Harr B (2000a). *Drosophila virilis* has long and highly polymorphic microsatellites. *Mol Biol Evol* **17**: 1641–1645.
- Schlötterer C, Zangerl B (1999). The use of imperfect microsatellites for DNA fingerprinting and population genetics. DNA profiling and DNA fingerprinting. In: Lubjuhn T (eds) Birkhäuser: Basel, pp 153–165.
- Spicer GS (1992). Reevaluation of the phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae). *Ann Entomology Society of America* **85**: 11–25.
- Spicer GS, Bell CD (2002). Molecular phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae) inferred from mitochondrial 12S and 16S ribosomal RNA genes. *Ann Entomol Am* **95**: 156–161.
- Throckmorton LH (1982). The *virilis* species group. In: Thompson JN (ed) *The Genetics and Biology of Drosophila*. Academic Press: New York, pp 227–296.
- Ting CT, Tsaur SC, Wu CI (2000). The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proc Natl Acad Sci USA* **97**: 5313–5316.
- Ting CT, Tsaur SC, Wu ML, Wu CI (1998). A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**: 1501–1504.
- Vieira CP, Coelho PA, Vieira J (2003). Inferences on the evolutionary history of the *Drosophila americana* polymorphic X/4 fusion from patterns of polymorphism at the X-linked paralytic and *elav* genes. *Genetics* **164**: 1459–1469.
- Vieira J, Charlesworth B (2000). Evidence for selection at the fused locus of *Drosophila virilis*. *Genetics* **155**: 1701–1709.
- Vieira J, McAllister BF, Charlesworth B (2001). Evidence for selection at the fused1 locus of *Drosophila americana*. *Genetics* **158**: 279–290.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>).